Proportions (z)

One sample sample is randomly selected from population of ____ ; sample size is < 10% of population size

sample size is large enough: **hyp test:** np ≥ 10 and n(1 – p) ≥ 10 **Conf Int:** $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

So I will use z procedures for a population proportion.

hyp test:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

conf interval:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Two samples both samples are randomly selected and selected independently

both sample sizes large enough $n_1\hat{p}_1 \geq 10$ $n_1\hat{q}_1 \geq 10$ $n_2\hat{p}_2 \geq 10$ $n_2\hat{q}_2 \geq 10$

So I will use z procedures for the difference of two population proportions.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p_c = \frac{TotalSuccesses}{TotalSamples} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Chi-Square (Counts)

$$\chi^2 = \Sigma \frac{(Obs - Exp)^2}{Exp}$$

	goodness-of-fit	homogeneity	independence
hypothesis	Is sample prop diff from hypothesized population distribution? H_O: they are same (be specific) H_A: they are different	proportions of 2 or more pop are different H_O: $p_1 = p_2 = p_3$ H_A: they are different	two variables are independent (no association) H_O: no association between …. H_A: an association between …
assumptions	all *expected* counts ≥ 5 random sample	all *expected* counts ≥ 5 random, **independent** samples	all *expected* counts ≥ 5 random sample
	So, I will use chi-square procedures for goodness of fit test with k-1 degrees of freedom (k = # of categories)	***So, I will use chi-square procedures for test for homogeneity with (r-1)(c-1) df***	***So, I will use chi-square procedures for test for independence with (r-1)(c-1) df***

MEANS (z or t)

One sample (st dev of population is <u>known</u>)
sample is randomly selected or assigned
population is fairly normally dist
OR sample size is large enough (n > 30)
OR boxplot of sample shows no outliers, so plausible that pop is normal
So, I will use z procedures for pop means

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

One sample (st dev of population is <u>NOT KNOWN</u>) <u>use t</u>
sample is randomly selected or assigned
population is fairly normally dist
OR sample size is large enough (n > 30)
OR boxplot of sample shows no outliers, so plausible that pop is normal
So, I will use t procedures for means with n-1 df

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \qquad \bar{x} \pm t * \frac{s}{\sqrt{n}}$$

Two samples Check **both** samples for:
samples are independent, random from____
st dev of populations are **known**
population is normal OR
sample sizes are large (n ≥ 30)
So, I will use z procedures for difference of 2 population means

Two samples: Check **both** samples for:
samples are random, <u>independent</u> (except treatments)
st dev of populations are **not known** **<u>use t</u>**
populations are normal OR
sample sizes are large (both n ≥ 30) OR boxplots of samples show no outliers, so plausible that both pops are normal

So, I will use t procedures for difference of 2 population means using smaller n-1 degrees of freedom

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}} \qquad (\bar{x}_1 - \bar{x}_2) \pm t * \sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}$$

<u>Matched pairs</u>
- matched (state how);
- sample randomly selected from pop of differences
- large (n ≥ 30) OR boxplot of <u>single</u> list of <u>*differences*</u> shows no outliers

(one list of differences is created from two <u>matched</u> lists)
So I will use <u>t</u> procedures for matched pairs (one sample of differences)
n-1 df formulas: see above-right for **<u>one sample t</u>**

t test for slope or model utility test or linear regression t test

hypothesis H_o: $\beta = 0$
H_a: $\beta \neq 0$

conditions
scatterplot of the data shows a linear pattern
variability of residuals does not appear to be changing with x
We hope that the residuals from the line at any given x are normallly distributed with mean = 0
(check histogram or normal probability plot of residuals)

$$t = \frac{b_1}{s_{b_1}} \qquad b_1 \pm t * s_{b_1}$$

So, I will use the linear regression t procedure with df = n - 2